Are You Really the Company You Keep? Reconciling Negative Peer Effects in College Achievement*

Ryan Brady

Department of Economics

United States Naval Academy

Michael Insler

Department of Economics

United States Naval Academy

Ahmed S. Rahman

Department of Economics

United States Naval Academy

December 2014

*Many thanks to participants of the Alan Stockman Conference at the University of Rochester in October 2014 for their helpful feedback.

Abstract

This paper examines the spillover effects on the academic performance of freshmen from their peers. Existing studies produce many contradictory findings. A rich dataset covering a 22-year history of the random assignment of students to close-knit peer groups at the U.S. Naval Academy, coupled with a lack of student discretion over freshman-year course choices, affords an opportunity to better understand peer effects in different social settings. We find negative peer effects at the broader company level and positive peer effects at the more narrow course-company level. We suggest that the marginal effect of peer characteristics may change sign because of differences in the underlying mechanism of peer influence.

Keywords: Peer effects, social network formation, academic achievement, homophily

JEL Codes: D85, I21, I23, I26, J24

1 Introduction

Economists have closely studied the role of educational peer effects in all levels of schooling but, so far, have had difficulty extending these insights into the policy arena. Research has demonstrated the importance of peer effects in higher education in particular. Studies of institutional data from Dartmouth College, Williams College, the University of Maryland, Berea College, the United States Military Academy (USMA), and the United States Air Force Academy (USAFA) have revealed peer effects of various sizes on a range of academic outcomes.[1] Some of most notable effects have been observed at USAFA, where Carrell et al. (2009) estimate that a 100-point increase in the peer-group average SAT verbal score increases freshman students' GPAs, on average, by 0.4 grade points on a 4.0 scale.[2] In a follow-up study, Carrell et al. (2013) analyze a direct intervention in which the researchers themselves allocated incoming students into peer groups designed to positively influence academic marks, as predicted by their historical estimates of peer effects. The intervention, however, backfired; the targeted beneficiaries of the experiment experienced small but statistically significant reductions in their grades. The implications of these findings in the peer effects literature are profound, as Carrell et al. (2013) directly state:

> Importantly, our results highlight both the significant role that peers play in the education production process and the theoretical difficulties in manipulating peers to achieve a desired policy outcome..... [Policy] interventions can affect patterns of endogenous social interaction. As such, we believe that endogenous responses to large policy interventions are a major obstacle to foreseeing the effects of manipulating peer groups for a desired social outcome. (Pg. 856-857)

In this paper, we attempt to unpack the potentially complex process of peer group formation and their effects. Studies such as Lyle (2007) and Carrell et al. (2009) rely on the random assignment of students (cadets at USMA and USAFA, respectively) into predetermined peer groups (hereafter referred to as "companies") to identify peer effects. This approach, however, implicitly assumes that students interact with all peers homogeneously, when it is more likely that the structure of social interactions is complicated. Specifically, students may form peer subgroups within companies, so that spillover effects differ across peers (see Manresa, 2013). Students may also differ in the degree of interaction with their peers, so that spillovers vary in intensity. Without further information regarding actual interactions among students, policy lessons are limited.

Students at the United States Naval Academy (USNA) have no discretion either over the assignment of their companies or their course enrollment during their first semester of freshman year. We exploit variation

[1] See Sacerdote (2001), Zimmerman (2003), Foster (2006), Stinebrickner and Stinebrickner (2006), Lyle (2007), and Carrell et al. (2009).

[2] The researchers suggest that their estimates are larger than previous findings because of the size (approximately 30 students) and critical role of peer groups at USAFA compared to more narrow roommate linkages seen in other studies.

in pre-treatment peer ability at the company level (i.e., the average of freshman companymates' SAT scores), as well as variation in pre-treatment peer ability at the company-course level, to analyze how spillovers may differ across various observable subgroups of students. Specifically, we compare spillovers generated by freshman students in the same companies *and* the same courses to those generated by the more broadly defined company. Unique to other peer effects studies, this allows us to measure peer effects within more narrowly defined subgroups that are clearly designated to participate in common tasks (i.e., studying similar coursework together, as opposed to living together while taking disparate courses).

Service academies' peer groups are well suited for studying peer effects because students spend an inordinate amount of time with their companymates in all aspects of college life. While previous studies have looked at peer effects on college students at the dorm-room level, dorm-floor level, and at the service academy group level (Carrell et al., 2009; Lyle, 2007), ours is the first to examine the course level within residential groups, at which there is a very natural opportunity for collaboration on the same tasks. Unlike previous studies, we are able to measure these narrower effects because we observe data on every USNA student, and their courses, who graduated between 1991 and 2012; our data contain more than 100,000 fall semester grades from over 20,000 freshmen.

The primary finding of our paper is twofold. First, we find *negative* peer effects at the company level. For STEM[3] courses, average peer ability across all freshman companymates, as measured by both verbal and math SAT scores, negatively affects own grades. This is a rare result in the peer effects literature that focuses on the college achievement; however, it is consistent with the findings of Carrell et al. (2013) in their natural experiment at USAFA. Our negative result is estimated from a standard, reduced-form specification, and it is robust to numerous sensitivity checks (which we discuss later). Second, we find that at the company-course level, average peer ability *positively* affects student performance, but only for relatively small companymate-coursemate peer groups. This contrasting result conforms with the idea that students may avoid interacting with differently-abled peers in broader social settings, while exhibiting closer interactions in smaller groups that perform specific common tasks.

Our paper helps explain some of the seemingly contradictory findings in the literature. Moreover, our emphasis on subgroup formation within broader peer groups stresses the need to understand the context of social settings and how common tasks can magnify endogenous peer effects. Failure to take these factors into consideration when attempting to manipulate peer groups can otherwise produce disastrous outcomes. We expound on these themes in the sections that follow.

The remainder of the paper is organized as follows. Section 2 provides a brief review of the key findings from the literature of most interest to our paper, and it also offers a theoretical motivation for our

[3]Refers to courses focused in science, technology, engineering, or mathematics.

empirical findings. In Section 3 we explain our setting and dataset in more detail. The sections thereafter summarize the estimation strategy and results, and provide discussion.

2 Motivation and Conceptual Framework

2.1 Identifying Peer Effects

The measurement of peer effects is complicated by "reflection." Peer effects are the reflection of our own image as it is cast upon our friends, who in turn cast their image back upon us.[4] This reflection is a function of our respective backgrounds, our current interaction with each other, and the environment in which we interact. Manski (1993) categorizes each of these factors in turn as contextual effects, endogenous effects, and correlated effects. Endogenous effects are a function of simultaneity, where peer group members affect each other and the observer cannot tell who is really affecting whom (see Glaser, 2009 for succinct discussion).[5] Contextual effects stem from the predetermined attributes (innate ability or training) of the person being affected or from the peer group doing the affecting. The term "contextual" comes from the sociology literature whereas economists would call it "exogenous."

Correlated effects refer to either common factors affecting a peer group or to the tendency of people to consort with what they perceive as like-minded individuals. Common factors could include the quality of the teacher in a particular class, the air conditioner not working for an entire semester, and so on. The desire to associate with similar people is the selection problem. An individual may gravitate towards those with characteristics they fancy in themselves and hence in others (early risers, tidy, a love of classical music, and so on). The selection characteristics are likely correlated with outcomes for the group and this makes it difficult to identify a peer effect since the features of an individual are correlated with why he or she is a part of that group to begin with (Lyle, 2007).

Consider the following structural form of the reflection issue where the researcher specifies the effect of peers on student i as follows:

$$GPA_{ict} = \beta_0 + \beta_1 \overline{GPA}_{jct} + \beta_2 \overline{PRE}_{jc,t-1} + \beta_3 PRE_{ic,t-1} + u_{ict} \tag{1}$$

GPA is the usual academic measure for grade point average, PRE is a measure of the exogenous "pre-treatment" ability the students bring with them to college (this is the "contextual" effect in Manski's, 1993,

[4] Manski (1993) deliberately uses the term "reflection" to evoke one looking in a mirror. An alien observer would not be able to tell if the image in the mirror was initiating movement, or if the person in front of the mirror was doing so.

[5] There tends to be some confusion in the literature with respect to Manski's terms. "Reflection" is sometimes listed as a fourth category. However, based on our reading of Manski (1993) and papers such as Glaser (2009) and Durlaf and Ionnides (2010), we believe our description to be consistent with Manksi (1993).

parlance). The subscript captures the fact that one is concerned with person i in peer group c in the current time period. The j denotes peer group members, where $j \neq i$. $\overline{GPA}$ and $\overline{PRE}$ are the averages of those peer group members' GPAs and pretreatment characteristics, respectively. The specification can be amended to control for various fixed effects (we discuss this more later with respect to our estimation). The parameter β_1 measures the endogenous peer effect while β_2 captures the contextual peer effect.

The inability to identify peer effects comes in three related forms. First the unobserved selection process implies the error term will be correlated with both the group pre-treatment variable and the group (time t) performance variable. Second, group performance will be correlated with the error term via the endogenous effect (the simultaneity of peer influence). And third, measurement of the group performance effect may be further biased by a common shock affecting all members of the group. The error term can expressed as,

$$u_{ict} = \varepsilon_{ict} + \theta_{ct} + v_{ict} \tag{2}$$

where the selection factor is ε_{ict}, the common factor is θ_{ct}, and v_{ict} is the usual random component through which simultaneity feedback may occur. Hence, in the structural equation both peer effect parameters β_1 and β_2 are likely to be biased.

The literature approaches the bias in different ways. A recent cadre of papers deals with the selection issue using "natural experiments" where peer groups are randomly assigned. Given the random assignment, the selection issue is mitigated. Relatively recent examples include Sacerdote (2001) and Lyle (2007), who estimate the structural equation (or some close version of it) shown in equation (1). Sacerdote (2001) estimates peer effects with random dorm room assignments of freshmen at Dartmouth. He regresses freshman's GPA on own level of ability (some "pre-treatment" metric that includes SAT scores and high school class rank) on said freshman's roommates' pre-treatment ability and on roommate GPA. In this case Sacerdote (2001) estimates an unbiased estimate of roommate's background; however the estimate on the effect of roommates' GPA will be biased. The key to identification for the former effect is the random assignment voids any possible correlation between roommate pre-treatment ability and other factors that may influence freshman i's GPA—parental pressure, non-roommate peers, and other components of the error term. As long as those factors are uncorrelated with roommates' background then one has a "clean" estimate of the peer effect from the roommates' pre-treatment ability.

Sacerdote (2001) ultimately finds that roommate pre-treatment ability is not a statistically significant explanatory factor for freshman GPA—so there is no contextual peer effect. However, roommate GPA is a statistically significant predictor of freshman GPA, but the result is obviously plagued by simultaneity bias. Sacerdote, however, suggests that bias notwithstanding, the statistically significant coefficient suggests that a positive peer effect is evident in the data. That is, Sacerdote claims, the biased coefficient represents *some* peer effect, even if the measurement is afflicted by simultaneity.

Lyle (2007) finds similar results to Sacerdote (2001) using random company assignment of freshmen at the United States Military Academy (USMA) to control for selection in the estimation of the structural equation above. Rather than being assigned by dorms or dorm floors, freshman at the USMA are randomly put into companies, which is a peer group of all grade levels (freshman, sophomore, junior, and senior). Lyle (2007) finds the endogenous peer effect is statistically significant, but the contextual effect is statistically insignificant. In his application, however, Lyle (2007) emphasizes that a common shock may explain much of the endogenous effect. Sacerdote (2001) recognizes the possibility of common shocks and imposes dorm-level fixed effects (which makes little difference for his results). Using common factors unique to USMA Lyle (2007)'s results suggest the endogenous peer effect may not be very large once specific controls for common shocks are introduced.

Overall, the evidence for peer effects in these papers relies on a biased estimate of the endogenous effect. In contrast, far from the banks of the Hudson River, Carrell et al. (2009) use random "squadron" assignment of freshmen at the United States Air Force Academy (USAFA) to search for peer effects in a reduced form model. Such a model, found in Carrell et al. (2009), Manski (1993), Zimmerman (2003), and others can be written,

$$GPA_{ict} = \left(\frac{\beta_0}{1-\beta_1}\right) + \beta_3 PRE_{ic,t-1} + \left(\frac{\beta_2 + \beta_1\beta_3}{1-\beta_1}\right)\overline{PRE}_{jc,t-1} + \varepsilon_{ict} \tag{3}$$

where the parameters map to the structural equation shown in equation (1).[6] With the randomly assigned peer groups at USAFA, Carrell et al. (2009) cite numerous notable results. Their measure of peer pre-treatment—verbal SAT scores—is statistically significant and positive at the squadron level, yet weak at the roommate level (and there is no effect from living close together). Peers' average verbal SAT score is statistically significant predictor of academic performance for math and science courses, but not for language or physical education courses. In contrast, peers' average math SAT score is not statistically significant for math and science courses.

If the effect were driven by some sort of "standard of work ethic" among the company, then the peer effect should matter in all courses. Differences in peer effects across course types suggests that the interaction of the students in those courses ("study partnerships" so to speak) is the mechanism driving the peer effect (a suggestion that we bolster in our paper). Carrell et al.'s (2009) results also underscore the challenge in

[6]For details with respect to the structural model and the reduced form in this literature see Manski (1993). For other examples of papers that estimate the reduced form see Guryan et al. (2009), Stinebrickner and Stinebrickner (2006), and Foster (2006).

estimating the peer effect from the reduced form—it is not clear what aspect of the reflection issue is driving the result, or what the mechanism might be that determines the size and sign of the parameter estimate. For example, in equation (3) the estimated coefficient of interest is comprised of each of the structural coefficients. In Carrell et al.'s (2009) reasoning, the peer SAT score is picking up some sort of quality that reflects the positive spillovers of interaction with peers. The interaction of group members matters, not merely the existence of certain "smart" people in the group whom the others might strive to emulate.

The result in Carrell et al. (2009) is consistent with numerous findings in studies on education, in particular, that the composition of the peer group matters. Zimmerman (2003) cites research on primary school students, whose performance increases with average classroom-IQ, but at a diminishing rate. Zimmerman (2003) suggests that that mixing students of varying ability (rather than segregating students) should generate higher aggregate learning. Indeed, Carrell et al. (2009) find that low ability students benefit from proximity to high ability peers (more than average ability students do).

With composition in mind, Carrell et al. (2013) take the rare step to implement an "optimal" peer assignment experiment at USAFA to harness the positive spillovers found in Carrell et al. (2009). The peer group assignments were meant to improve the performance of the lowest ability students, where the primary treatment group ("bimodal squadrons") include low ability students alongside a larger fraction of peers with high SAT verbal scores. A second treatment group ("homogeneous squadrons") is comprised of primarily "middle ability" students. The control group is formed in the same manner as always at USAFA (see Carrell et al., 2013, for details).

Contrary to most evidence from the empirical literature, Carrell et al. (2013) estimate a *negative* treatment effect for the main treatment group of interest—the low ability students performed worse than similar students in the control group. Conversely, students in the homogenous treatment group performed better than their counterparts in the control group.

The authors suggest that homophily explains the results. The low-ability students in the treatment group were more likely to study with—and identify as friends—other low-ability students in their squadron (friends were identified via a follow up survey conducted by the authors). The higher ability students in the treatment squadron segregated themselves similarly. The middle ability students, on the other hand, benefited from the lack of interaction with low ability students. In the control group the tendency for such sorting was less apparent.

Foreshadowing Carrell et al. (2013), Zimmerman (2003) finds a positive effect from peers' verbal SAT for students in the middle 70 percent of the SAT distribution. But Zimmerman (2003) also finds no significant peer effects acting on students in the bottom or top 15th percentiles. Another example of a negative estimate (a rare occurrence in the literature) is from Foster (2006). For a sample of University of

Maryland students, she estimates a statistically significant negative peer effect on male students that stems from their peers' median SAT scores (but not from peers' average SAT scores).

2.2 Negative Peer Effects and Homophily

The negative peer effect found in the natural experiment instituted at USAFA may be evidence of the preponderance of homophily within a group. Carrell et al. (2013) note that the "bimodal squadron" treatment group in the USAFA experiment contained an unusually high number of high-ability types and an unusually high number of low-ability types (relative to the control group). This composition effect may have inspired a change in the formation of peer groups within the squadron in a different way than occurs within the randomly assigned control group (which includes a mixture of all ability-types).

The lesson from the USAFA experiment appears to be that negative peer effects are a function of the tendency for sorting within a group, and that there may be some tipping point with respect to group composition that leads a student in one direction (towards persons of the same characteristics) or the other. Unfortunately, the nature of this sort of endogenous group formation is a "black box," as Carrell et al. (2013) note towards the end of their paper. What can be conjectured, however, is how the tendency towards homophily might lead to the negative outcomes for lower-ability students.

2.3 A Simple Framework

To make more concrete some of the ideas mentioned above, let us consider a student who cares about only two things, his or her grades, defined as G, and his or her "homophily index," defined as H. Assume each of these depend on the student's own innate characteristics, and *potentially* the characteristics of his or her peers. Let us further assume that if the student interacts with the peer group, the student's homophily index is lower the more *different* he or she is from the group. On the other hand, with social interaction, the student's grades improve the *weaker* his or her characteristics are relative to the peer group. More specifically, assume that the student aims to maximize U, where

$$U = G^{\alpha} H^{1-\alpha} \tag{4}$$

Given a certain peer group, the student chooses either to interact with his or her peers (call this being "open"), or not interact (call this being "closed"), in order to maximize (4). Let us consider *ability* as the sole characteristic that matters, and that ability is innate. a_0 is the student's ability, while a_i is peer i's ability. We will assume in the exposition below that the student's peers all have at least the same ability as the student. Let us then have following functional forms:

$$G = \begin{cases} a_0^{\beta} & \text{if } closed \\ \left(a_0 + \gamma \sum_{i=1}^{n} (a_i - a_0) \right)^{\beta} & \text{if } open \end{cases} \tag{5}$$

$$H = \begin{cases} a_0 & \text{if } closed \\ a_0 - \sum_{i=1}^{n} (a_i - a_0) & \text{if } open \end{cases} \tag{6}$$

where n is the number of the student's potential peers, $\gamma > 0$ measures the degree of ability spillover to the student's grade performance, and $0 < \beta < 1$. The meaning of each measure is straight-forward. Equation (5) suggests interacting with peers of stronger ability helps the student achieve better grades, although at a diminishing rate. If the student chooses not to interact, the student relies solely on his or her own ability a_0. On the other hand, Equation (6) suggests interacting with peers of stronger ability reduces one's homophily index, thus lowering overall utility. The student's decision to interact with his or her peer group is essentially based on which effect dominates.

2.4 Numerical Examples

Let us demonstrate the simple framework above with some numbers. Consider an average student with $a_0 = 3$. She has a potential peer who is either also of average ability ($a_1 = 3$) or high ability ($a_1 = 4$). Assume that $\alpha = \beta = 0.5$, and that $\gamma = 5$. In this case, it is straight-forward to see that she would be indifferent to being open or closed if the peer was of average ability ($U = 3^{0.25} \times 3^{0.5} = 2.28$ in either case), but she would be open if her peer was of high ability ($U_{open} = 2.37; U_{closed} = 2.28$). In this case we can safely say that higher peer ability will benefit the student's grade performance.

Now let us consider the case with two potential peers, each of who could be of average ability or high ability. There are thus three possible cases. With two average ability peers, the student is again indifferent between being open or closed. With one high ability peer and one average peer, the student would choose to be open ($U_{closed} = 2.28 < U_{open} = 2.37$). But with two high ability peers, the student would choose to be closed ($U_{closed} = 2.28 > U_{open} = 1.90$). In this case we observe a non-monotonic relationship between grade performance and average peer ability - as peer ability rises past some point, grade performance actually declines. The reason for this is straight-forward. Peer ability raises a student's grades at a declining rate, while the homophily index declines at a constant rate. At a certain point the dis-utility experienced outweighs the gains from grades, and the student chooses not to interact.

There are a couple of lessons from this simple set-up.

Proposition 1. *For a given distribution of peer ability, there exists some size of peer group $\hat{n}$ such that a* *marginal increase in peer size from $\hat{n}$ will be associated with lower grades for the student.*

The intuition of this is simple. With larger peer groups the open student will be increasingly exposed to peers of different ability. Because peers' positive influence on grades face diminishing returns, at some size the homophily costs of being open outweigh the gains from grades. In reality of course students need not be totally closed or open, but instead choose an openness "range" where they choose the size and composition for their peers. But the idea here still supports this more general case. Students who become more closed will choose peers of more similar ability, and this will cause grade deterioration.

Proposition 2. *For a given distribution of peer ability and size of the peer group, there exists some $\hat{\gamma}$ such that if $\gamma > \hat{\gamma}$, a marginal increase in average peer ability will be associated with better grades for the student.*

This simply means that the positive peer spillover needs to be sufficiently large in order for students to be willing to always be open. If $\gamma < \hat{\gamma}$ on the other hand, the peer effects on individual grades is unclear. If stronger peer characteristics induce students to become more closed off to their peers, the opposite can occur.

This very simple framework demonstrates that endogenous peer group formation can be important for peer effects even when groups of peers appear to be exogenously created. The framework can be easily extended to incorporate cases where students choose an optimal degree of openness on the *intensive* margin. In this case it can be shown that for γ smaller than some threshold, rising average group quality causes low-ability students to choose a smaller *set* of like-ability peers, thereby causing the students' grades to fall. The implication of this more general case is the same—rising average quality of the group can lower student performance in certain social settings.

The next sections test some of the implications presented above with evidence from freshmen enrolled at the U.S. Naval Academy.

3 Our Setting

3.1 Random Assignment

The United States Naval Academy (USNA) provides an ideal setting to identify the effect of social interactions on academic achievement. Upon arrival, every freshman is assigned into a company. All students live in one on-campus dormitory, which houses 30 companies of approximately 150 students, each containing an even mix of freshmen, sophomores, juniors, and seniors. A student's company is his or her primary peer group. The company assignment procedure, which is administered by the Admissions Office, is designed to produce a diverse but randomly allocated mix of students in each company. Students are first randomly spread across companies based on predetermined characteristics: race, gender, home state, recruited athlete status, prior military service, and attendance at a one year Naval Academy preparatory school. After these initial stratifications, administrators randomly assign all remaining students to companies.[7] The key features of the procedure are that students have no control over the outcome—for instance, USNA does not solicit interests, lifestyle details, or roommate preferences as is typical at other universities—and it produces an allocation that is effectively random. For instance, the mechanism prevents students from sorting into residence hall groupings that could offer academic advantages.

3.2 Data

Our data, which were compiled with the aid of USNA's Office of Institutional Research, contain far more observations—more than 100,000 grades from over 20,000 first-semester freshmen-–than comparable studies from other institutions.[8] In addition to every grade assigned to freshmen in the classes of 1991-2012, we observe the following student specific characteristics: race/ethnicity, whether a recruited athlete, whether previously enlisted in the armed forces, whether attended a one-year preparatory school prior to enrolling at USNA, and math and verbal SAT scores. Table 1 contains summary statistics for grades obtained by freshmen during their initial fall semester. USNA is predominantly male and white, although the institution has become more diverse in recent years. SAT scores are high; the sample averages of math and verbal SAT scores are 661 and 638, respectively.

In addition to the random assignment of peer groups, there are other features that make USNA an ideal laboratory in which to examine peer effects. Freshman students at USNA have no ability to select their courses, and they have virtually no ability to choose specific course sections (e.g., to select an instructor with a "kinder" reputation). All freshmen must pass or validate a set of 11 core courses in a range of subject areas such as calculus, chemistry, political science, and naval history. These courses form their entire first year schedule, with few exceptions.[9] Additionally, USNA has relatively low grade inflation. Based on our sample, average GPAs have risen over the years, from an average of 2.7 in 1991 to 3.0 in 2009. Rojstaczer and Healy (2010) show that average GPAs from a large sample of American four-year colleges have increased

[7]This procedure is very similar to USAFA's, described by Carrell et al. (2009). It differs from the procedure at the United States Military Academy (USMA), described by Lyle (2007), which additionally produces an even mix of academic ability (proxied by incoming SAT score) across companies. Lower variation in academic achievement at USMA across companies yield an environment in which it is much more difficult to estimate peer effects.

[8]For comparison, Carrell et al. (2009) utilize a sample of approximately 20,000 grades.

[9]If a student expressed interest in the majoring in a critical language (i.e., Chinese or Arabic), then the student would take courses in that language starting freshman year and, therefore, postpone a few mandatory classes until their sophomore or junior year. Students may also test out of one or more core classes, which would lead them to enroll in higher level classes during their freshman year. For example, a student with sufficient prior experience in calculus could enroll in Calculus II during fall of freshman year while his peers enroll in Calculus I.

approximately linearly from 2.9 in 1991 to 3.1 in 2006. Thus the trend at USNA is steeper but averages remain lower, compared to other schools. Given that GPAs are far more frequently bounded above than below, this suggests both that there is higher grade variation at USNA and that grades produce higher signal-to-noise ratios than at other institutions.

4 Econometric Model

4.1 Baseline Model

We envision a freshman's first semester academic grades following a "production process" with inputs: (1) own high school (i.e., pre-USNA) characteristics; (2) peers' high school characteristics; (3) variation that is specific to course and academic year. Given the similarities between USNA and USAFA, we adopt a specification very similar to that of Carrell et al. (2009); we use the following linear model:

$$G_{igct} = \alpha + \beta Z_{igt} + \gamma \frac{\sum_{k \neq i} Z_{kgt}}{n_{gt} - 1} + \delta X_{igt} + \theta Y_{ct} + \eta_t + \varepsilon_{igct} \tag{7}$$

In this specification, G_{igct} is grade (on a standard four point scale) of student i in peer group g (i.e., company g) for course c in academic year t.[10] We only use grades from fall semester of freshman year to avoid issues related to self-selection into courses in subsequent semesters. Z_{igt} is student i's pre-USNA characteristics that may directly affect academic achievement, for which we proxy with SAT math and SAT verbal scores. Z_{kgt} includes SAT math and verbal scores for student k, who is one of i's companymates. γ captures the influence of i's "average" companymate, as we average over each of these characteristics for all $k \neq i$. X_{igt} is the set of controls for student i that enter the stratified company assignment procedure: race/ethnicity, gender, whether a recruited athlete, whether attended a feeder school, and whether possessed prior military experience. Y_{ct} includes information that is specific to course-years (in a robustness check, we use the one-year lag of average course grade as a proxy for Y_{ct}) and η_t is a set of academic year dummies. ε_{igct} represents all omitted factors. By construction, ε_{igct} cannot be correlated with Z_{kgt} because peer assignment is random and all peer characteristics in Z_{kgt} were established before arriving at USNA (i.e., prior to treatment).

4.2 Incorporating Coursemate Subgroups into the Model

For the model in Equation (7), there remains an interpretation issue; within γ, we cannot identify the relative makeup of contextual effects from peer attributes or endogenous effects from contemporaneous

[10] Academic year t does not represent the time period in the usual context of panel data, because we use each student's grades only from his or her initial semester. In other words, there is only one time period t for each student i, but we include the subscript to indicate possible academic year effects (e.g., grade inflation).

peer achievement. Given the discussion in Section 2, it is useful to recall the endogenous and contextual components that comprise our estimate of the peer effect in Equation (7). Broadly written, the peer effect is:

$$\gamma = \left(\frac{\beta_2 + \beta_1 \beta_3}{1 - \beta_1} \right) \tag{8}$$

Our conceptual framework in Section 2 indicates that negative peer effects may arise when students choose to group themselves by ability within a broader group, and this effect becomes more pronounced as the group size increases. In that setting, does the negative peer effect arise due to the simultaneous peer interaction (the endogenous effect) or from the contextual effects?

First, consider the endogenous effect captured by β_1. For β_1 to be negative, it must be that the high performance of the peer group somehow leads to individual students' performance, on average, being harmed, distinct from the contextual differences. Instead, it seems highly likely that any idiosyncratic negative effects would be dominated by other, positive contemporaneous effects. It is also reasonable to assume that $\beta_1 < 1$: Peers' contemporaneous achievement should not have a multiplied impact on own achievement. Therefore, we assume:

Proposition 3. $0 \leq \beta_1 < 1$

Second, consider the coefficient on own ability, β_3. It is very reasonable to assume that a student's own predetermined ability is positively correlated with his or her grades:

Proposition 4. $\beta_3 > 0$

A policy maker designing optimal peer groups only has control over each group's composition of pre-treatment characteristics. That is, the policy maker's influence works only through β_2. But given the natural experiment at USNA, we can, at best, consistently estimate γ. Under Propositions 3 and 4, γ's sign is determined by β_2's sign, as well as β_2's magnitude relative to $\beta_1 \times \beta_3$. Thus, a negative estimate of γ could only stem from a strongly negative β_2. On the other hand, a positive estimate of γ could be produced by either a positive β_2, or a negative β_2 that is drowned out by relatively strong β_1 and β_3 parameters. While we cannot directly observe these latent parameters, our conceptual framework suggests that the contextual effect β_2 can conceivably be positive or negative. Therefore, positive estimates of γ in some peer group settings alongside negative estimates of γ in other settings must be, respectively, produced by a positive β_2 (or at least a drowned-out, negative β_2) and a negative β_2.

Exploiting USNA's joint random assignment into companies and course sections, we are able to test whether different sets of peers may influence academic achievement in different ways. In particular, we examine what we view as the most likely channel through which peer effects may operate: collaboration between students in the same company *and* course. We modify Equation (7) as follows:

$$G_{igct,s} = \alpha_s + \beta_s Z_{igt,s} + \gamma_s \frac{\sum_{k \neq i} Z_{kgct,s}}{n_{gct,s} - 1} + \delta_s X_{igt,s} + \theta_s Y_{ct,s} + \eta_{t,s} + \varepsilon_{igct,s} \tag{9}$$

In this model, the peer effect now stems from companymate-coursemates, rather than just companymates. To reflect that the size and direction of peer effects may differ depending on the size of one's peer group, Equation (9) also introduce peer group sizes, s. We estimate Equation (9) on subsamples of the data that are stratified by the sizes of companymate-coursemate peer groups. Figure 1 shows the distributions of these peer groups' sizes for humanities/social science and math/science grades, separately. Given the distributions, it is reasonable to use peer group stratifications between 2-10, 11-30, and over 30, for s.

Based on the discussion in Section 2, our prior is that $\gamma < \gamma_s$ for smaller peer groups s. By isolating smaller peer groups whose students are engaged in specific tasks, the background characteristics of their peers would conceivably matter more for their grade performance. In the conceptual framework, this would be consistent with a higher γ; students of disparate backgrounds would be more likely to interact and thereby benefit from higher-ability peers.

5 Results

5.1 Baseline Model

Table 2 contains OLS estimates of Equation (7). We cluster standard errors by academic year-company groups and estimate separate regressions for math/science and humanities/social science course grades. We also estimate models that include the standard deviation of companymates' SAT scores, to check if company-related peer effects act through the heterogeneity of peer groups.

Estimates of all control variables have reasonable signs and magnitudes. Women perform better than men in humanities and social sciences, but worse in math and science courses. Minorities, recruited athletes, and preparatory school attendees tend to under-perform in the classroom, likely due to selection. Freshmen who were previously enlisted in the armed forces earn math and science grades that are higher, on average, by 0.16 grade points. Own SAT scores are positively associated with own grades. A 100-point increase in verbal SAT score is associated with nearly 0.3 additional grade points in humanities/social science courses, and a 100-point increase in math SAT score is associated with a 0.5 grade point increase in math grades. Verbal SAT score's association with humanities/social science grades is an order of magnitude larger than its association with math/science grades, and likewise for math SAT score.

For humanities and social science courses, estimates of peer effects are insignificant. Noting the large sample of 47,536 grades assigned to freshmen in the fall semester, there is no evidence that the average or standard deviation of SAT scores affect grade output. For math and science courses, peer effects that act through SAT scores are *negative* and significant at the one percent level, but very small. A one standard deviation (9.9 point) increase in companymates' average SAT math score yields a 0.02 point decrease in own math and science course grades, and a one standard deviation (12 point) increase in companymates' average SAT verbal score yields a 0.019 point decrease. We find no significant effects for the standard deviation of companymates' SAT scores, indicating that simple heterogeneity in companymates' quality does not produce results. This contrasts Lyle (2007), perhaps due in part to the fact that we can utilize roughly seven times more observations for each specification.

Are we to believe that interacting with peers with strong academic ability causes one to perform more poorly in school? Far more likely, students are self-selecting in particular ways within companies that result in these negative grade effects. The conceptual framework in Section 2 indicates that student performance can suffer with higher peer ability when grade spillovers are weak (this is where $\gamma < \gamma_s$). Here we can argue that in a context where students live together but potentially take different courses, grade spillovers would indeed be weak, leading to these results.

5.2 Course Subgroup Model

Table 3 contains OLS estimates of Equation (9). Standard errors are clustered by academic year-company groups. We now stratify by both course type (humanities/social science, math/science) and peer group size (2-10, 11-30, 31 or more), where freshman i's peer group—and there is potentially a different peer group for each of i's distinct grades—is defined by the set of freshmen companymates contemporaneously taking the same course.

Across all models, estimates of control variables' coefficients are consistent with previous findings. Striking differences appear when comparing the coursemate-specific peer effects across different peer group sizes. For small peer sizes (2-10), we observe positive peer effects. For humanities and social science courses, small peer groups with stronger *verbal* SAT scores positively affect individual grades; for math and science courses, small peer groups with stronger *math* SAT scores positively affect individual grades. These estimates are significant at the 0.1 percent level and comparable in size to estimates in previous literature: A one standard deviation (11.7 point) increase in coursemate-companymates' average SAT verbal score yields, on average, a 0.013 point higher humanities/social science course grade, and a one standard deviation (10.0 point) increase in coursemate-companymates' average SAT math score yields, on average, a 0.04 point higher math/science course grades. Carrell et al. (2009) find that a one standard deviation increase in peer verbal

SAT results, very comparably, in 0.05 additional grade points. On the other hand, Zimmerman (2003) finds a much smaller but still positive peer effect for roommates at Williams College. Our positive peer effects disappear for larger peer groups. There is modest evidence of a negative peer effect through math SAT for groups of size 11-30, and for peer groups greater than thirty, the group compositions approach the full company of freshman students (i.e., where all students in the company take the same course).

Results are consistent with our framework discussed in Section 2. Smaller peer groups working on similar tasks are far more likely to interact with each other and influence one another's performance (in the framework this would be consistent with a larger γ). These findings suggest that endogenous peer group effects exist and are strong.

5.3 Robustness Checks

In this subsection, we test the robustness of the findings in Table 3. First, we use a placebo peer group: for each grade G_{igct}, we now define freshman i's peer group as his or her company g mates who are *not* taking course c. We stratify grades, as always, by humanities/social science courses and math/science courses. Table 4 displays results. The negative peer effects suggested in Table 2 are echoed here. These results bolster the previous finding that students in the same course generate a strong effect for positive peer spillovers.

A potential concern is that structural differences in courses produces the differences in peer effects across peer group sizes seen in Table 3. For example, grades assigned when coursemate-companymate peer groups are only of size 2-10 may more frequently come from courses with lower enrollment. If it were the case that, in such courses, instructors tend to assign higher grades, then our positive estimates of γ_s would be biased upward by course size-related grade inflation. Therefore, in the second robustness check, shown in Table 5, we re-do the regressions of Equation (9) in Table 3, but now we include a measure of the average grade students received in that course the previous year (a proxy for Y_{ct}).[11] Conceptually this may be an important control. Estimates show, however, that our positive peer effects for small peer group sizes hold up with the additional control. Freshman courses associated with smaller peer groups do not appear to distribute higher or lower grades in any systematic way. We also see a fair amount of grade persistence over time—the average grade of a course in prior years is a strong predictor of one's own grade in that course. But the addition of this control keeps our overall findings on peer effects at the company-course level intact.

[11]Technically, the variable "Avg. course grade (previous year)" is the residual from a regression of the one-year lag of the course's average grade on the peer group average SAT scores of students enrolled in the course. We include this measure to capture those factors other than student ability that may influence faculty to give higher or lower course grades from the school average.

6 Conclusion

This paper attempts to reconcile the seemingly inconsistent findings of positive peer effect spillovers in some contexts and seemingly negative spillovers in other contexts. In short, our paper highlights the idea that context matters. In large social settings or living arrangements, more favorable average peer attributes can perversely lower individual performance as individuals increasingly group with those of like-traits. In other settings where individuals are engaged in common work on tasks, such grouping patterns can be overcome.

Our results can guide policy interventions meant to harness peer effects. Cognizance about potential complications of peer group assignment is critical. Increasing peer quality can foster beneficial collaboration in common tasks in close-knit, small peer groups, but the specific design of "optimal peer groups" for positive educational outcomes is indeed complicated and remains an avenue for future research.

References

[1] Carrell, S., Fullerton, R., and West, J. (2009). Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics,* 27(3), 439-464.

[2] Carrell, S., Sacerdote, B., and West, J. (2013). From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica,* 81(3), 855-882.

[3] Durlaf, S., and Ioannides, Y. (2010). Social Interactions. *The Annual Review of Economics,* 2, 541-578.

[4] Foster, G. (2006). It's Not Your Peers, and It's Not Your Friends: Some Progress Toward Understanding the Educational Peer Effect Mechanism. *Journal of Public Economics,* 90, 1455-1475.

[5] Glaser, D. (2009). Teenage Dropouts and Drug Use: Does the Specification of Peer Group Structure Matter? *Economics of Education Review,* 28, 497-504.

[6] Guryan, J., Kroft, K., and Notowidigdo, M. (2009). Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments. *American Economic Journal: Applied Economics,* 1(4), 34-68.

[7] Lyle, D. (2007). Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point. *Review of Economics and Statistics,* 89(2), 289-299.

[8] Manresa, E. (2013). Estimating the Structure of Social Interactions Using Panel Data. Working Paper.

[9] Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies,* 60, 531-542.

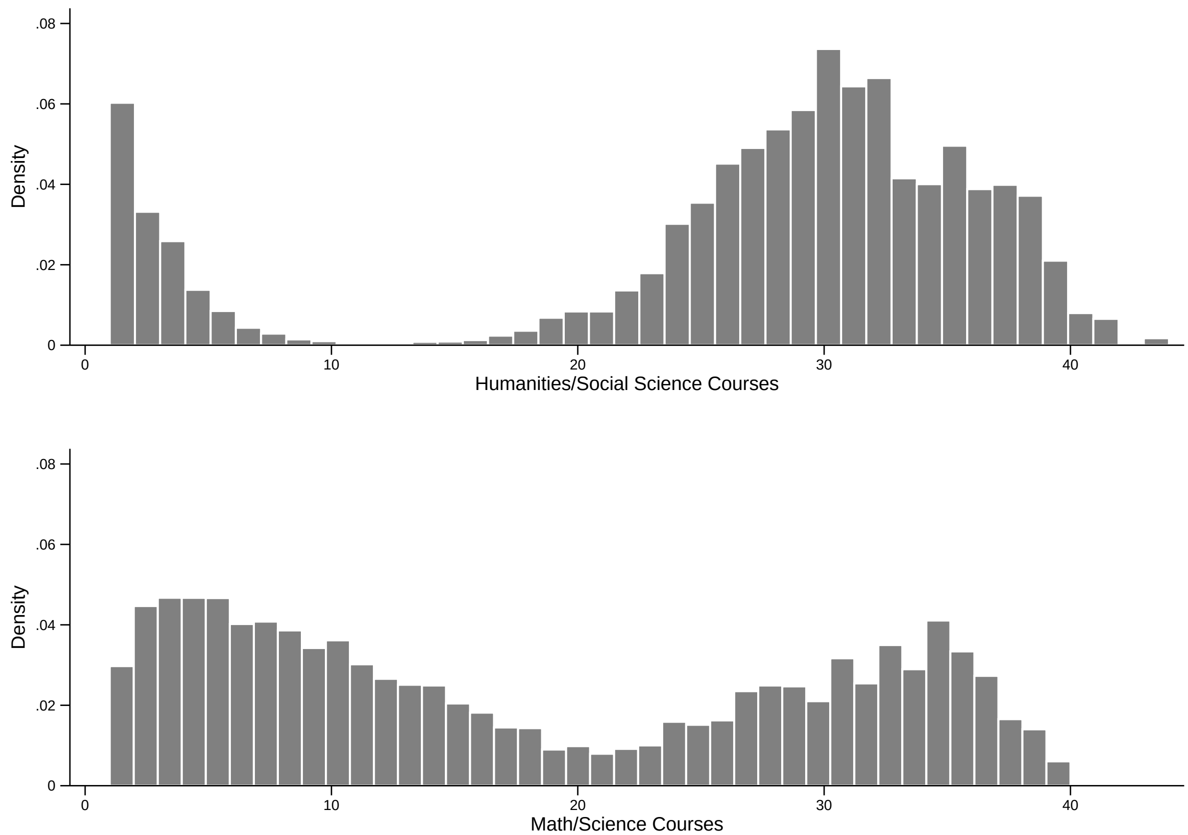
[10] Rojstaczer, S. and Healy, C. (2010). Grading in American Colleges and Universities. Teachers College Record. March 4, 2010.

[11] Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics,* 116(2), 681-704.

[12] Stinebrickner, R. and Stinebrickner, T. (2006). What can be Learned about Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds. *Journal of Public Economics,* 90, 1435-1454.

[13] Zimmerman, D. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics,* 85(1), 9-23.

Figure 1: Distributions of Coursemate-Companymate Peer Group Sizes



Note: Plots show histograms of coursemate-companymate peer group sizes, as peer groups are defined in Equation (9). Data are stratified by course type.

Table 1: Summary Statistics

Variable	Mean*	St. Dev.
Female	0.15	0.36
Race/ethnicity:		
Black	0.06	0.24
Asian	0.04	0.19
Hispanic	0.08	0.27
White	0.79	0.41
Other	0.02	0.15
Recruited athlete	0.27	0.44
Prior enlisted	0.09	0.29
Feeder source:		
NAPS	0.17	0.38
Foundation school	0.07	0.25
Direct Entry	0.75	0.44
Other	0.02	0.13
Own SAT math	661	64.2
Own SAT verbal	638	68.7
Peer SAT math (company average)	662	9.9
Peer SAT verbal (company average)	638	12.0
Number of Observations	100,146	

Note: *Column shows sample means for SAT scores and sample proportions for all other variables.

Table 2: Regressions - Freshman Companymates as Peers

Dependent Variable: Grade	Hum/SS	Hum/SS	Math/Sci	Math/Sci
Female	0.0436***	0.0436***	-0.0888***	-0.0888***
	(0.0114)	(0.0114)	(0.0144)	(0.0144)
Race/ethnicity (ref.: White and other):				
Black	-0.276***	-0.276***	-0.267***	-0.267***
	(0.0176)	(0.0176)	(0.0213)	(0.0213)
Asian	-0.0916***	-0.0915***	-0.0796***	-0.0796***
	(0.0196)	(0.0196)	(0.0237)	(0.0237)
Hispanic	-0.144***	-0.144***	-0.200***	-0.200***
	(0.0155)	(0.0155)	(0.0191)	(0.0192)
Recruited athlete	-0.0826***	-0.0826***	-0.132***	-0.132***
	(0.00934)	(0.00935)	(0.0118)	(0.0119)
Prior enlisted	0.00906	0.00893	0.160***	0.160***
	(0.0173)	(0.0172)	(0.0208)	(0.0208)
Feeder source (ref.: Direct entry):				
NAPS	-0.162***	-0.162***	0.140***	0.140***
	(0.0128)	(0.0128)	(0.0159)	(0.0159)
Foundation school	-0.0593***	-0.0592***	-0.0827***	-0.0827***
	(0.0156)	(0.0156)	(0.0208)	(0.0209)
Other	-0.0281	-0.0280	0.247***	0.247***
	(0.0395)	(0.0395)	(0.0417)	(0.0416)
Own SAT math	0.000532***	0.000531***	0.00510***	0.00510***
	(0.0000770)	(0.0000770)	(0.000101)	(0.000101)
Own SAT verbal	0.00298***	0.00298***	0.000295***	0.000296***
	(0.0000721)	(0.0000721)	(0.0000840)	(0.0000840)
Peers Effects:				
Peers' SAT math (avg.)	-0.000906	-0.000924	-0.00198**	-0.00198**
	(0.000550)	(0.000546)	(0.000709)	(0.000709)
Peers' SAT math (std. dev.)		-0.00198		
		(0.00222)		
Peers' SAT verbal (avg.)	-0.000784	-0.000787	-0.00158**	-0.00156**
	(0.000408)	(0.000407)	(0.000561)	(0.000563)
Peers' SAT verbal (std. dev.)				0.000822
				(0.00220)
Constant	1.605***	1.746***	1.155*	1.087*
	(0.408)	(0.418)	(0.498)	(0.513)
Number of Observations	47,536	47,536	52,610	52,610
R^2	0.168	0.168	0.144	0.144

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. OLS estimations are carried out separately for humanities and social science course grades (column 1) and math and science course grades (column 2). Coefficients for academic year dummy variables are included in the estimations but are not shown. Standard errors are clustered by company-academic year groups.

Table 3: Regressions - Freshman Coursemate-Companymates as Peers

Dependent Variable: Grade	Hum/SS	Hum/SS	Hum/SS	Math/Sci	Math/Sci	Math/Sci
Coursemate-peer group size:	2-10	11-30	31 or more	2-10	11-30	31 or more
Female	0.149***	0.0982***	-0.0487**	-0.0221	-0.142***	-0.102***
	(0.0243)	(0.0161)	(0.0178)	(0.0191)	(0.0231)	(0.0206)
Race/ethnicity (ref.: White and other):						
Black	-0.222***	-0.260***	-0.320***	-0.217***	-0.336***	-0.268***
	(0.0368)	(0.0260)	(0.0256)	(0.0313)	(0.0292)	(0.0330)
Asian	-0.0524	-0.0909***	-0.113***	-0.0165	-0.129***	-0.101**
	(0.0471)	(0.0265)	(0.0298)	(0.0341)	(0.0345)	(0.0384)
Hispanic	-0.0769*	-0.128***	-0.178***	-0.172***	-0.216***	-0.209***
	(0.0356)	(0.0214)	(0.0227)	(0.0286)	(0.0279)	(0.0284)
Recruited athlete	-0.111***	-0.0666***	-0.0920***	-0.0754***	-0.154***	-0.187***
	(0.0207)	(0.0127)	(0.0137)	(0.0179)	(0.0169)	(0.0183)
Prior enlisted	0.0511	-0.00567	0.0131	0.148***	0.180***	0.122**
	(0.0383)	(0.0220)	(0.0282)	(0.0311)	(0.0274)	(0.0378)
Feeder source (ref.: Direct entry):						
NAPS	-0.122***	-0.158***	-0.193***	0.0383	0.133***	0.272***
	(0.0304)	(0.0173)	(0.0198)	(0.0256)	(0.0227)	(0.0257)
Foundation school	-0.0746	-0.0567**	-0.0490*	-0.122***	-0.103***	0.0205
	(0.0387)	(0.0210)	(0.0227)	(0.0299)	(0.0295)	(0.0315)
Other	-0.0546	-0.00640	-0.0351	0.267***	0.129*	0.394***
	(0.0915)	(0.0558)	(0.0528)	(0.0646)	(0.0636)	(0.0602)
Own SAT math	0.000594***	0.000378***	0.000610***	0.00396***	0.00503***	0.00613***
	(0.000179)	(0.000107)	(0.000113)	(0.000150)	(0.000169)	(0.000145)
Own SAT verbal	0.00259***	0.00282***	0.00286***	-0.000352**	0.000389**	0.000945***
	(0.000163)	(0.000107)	(0.000104)	(0.000119)	(0.000130)	(0.000126)
Peer Effects:						
Coursemate-peers' SAT math (avg.)	-0.0000143	-0.00131*	-0.000195	0.000772***	-0.00260***	-0.000680
	(0.000215)	(0.000634)	(0.000787)	(0.000184)	(0.000575)	(0.000907)
Coursemate-peers' SAT verbal (avg.)	0.00111***	-0.000748	-0.000665	0.000135	-0.000818	-0.00172*
	(0.000186)	(0.000506)	(0.000651)	(0.000193)	(0.000615)	(0.000800)
Constant	0.0180	2.006***	1.105	-0.643***	1.091**	-0.897
	(0.156)	(0.420)	(0.593)	(0.132)	(0.383)	(0.700)
Number of Observations	6,183	20,559	20,794	19,801	19,101	13,708
R^2	0.296	0.139	0.164	0.146	0.143	0.213

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. OLS estimations are carried out separately for humanities and social science course grades (columns 1-3) and math and science course grades (columns 4-6). Estimations are further stratified by the size of the coursemate-peer group associated with each grade observation (sizes 2-10, 11-30, and over 30). Coefficients for academic year dummy variables are included in the estimations but are not shown. Standard errors are clustered by company-academic year groups.

Table 4: Robustness - Freshman Companymates Not in Same Course as Peers

Dependent Variable: Grade	Hum/SS	Math/Sci
Female	0.0655***	-0.0858***
	(0.0117)	(0.0145)
Race/ethnicity (ref.: White and other)		
Black	-0.268***	-0.271***
	(0.0183)	(0.0215)
Asian	-0.0821***	-0.0762**
	(0.0203)	(0.0235)
Hispanic	-0.139***	-0.202***
	(0.0158)	(0.0192)
Recruited athlete	-0.0821***	-0.130***
	(0.00963)	(0.0119)
Prior enlisted	0.00506	0.162***
	(0.0179)	(0.0209)
Feeder source (ref.: Direct entry)		
NAPS	-0.168***	0.131***
	(0.0133)	(0.0161)
Foundation school	-0.0564***	-0.0884***
	(0.0164)	(0.0210)
Other	-0.00336	0.250***
	(0.0408)	(0.0421)
Own SAT math	0.000531***	0.00493***
	(0.0000795)	(0.000102)
Own SAT verbal	0.00290***	0.000238**
	(0.0000734)	(0.0000848)
Peer Effects:		
Non-coursemate-peers' SAT math (avg.)	-0.000119	-0.00196***
	(0.000147)	(0.000252)
Non-coursemate-peers' SAT verbal (avg.)	-0.000185	-0.000744**
	(0.000109)	(0.000264)
Constant	0.752***	0.745***
	(0.108)	(0.178)
Number of Observations	42,740	51,670
R^2	0.166	0.147

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. OLS estimations are carried out separately for humanities and social science course grades (column 1) and math and science course grades (column 2). Coefficients for academic year dummy variables are included in the estimations but are not shown. Standard errors are clustered by company-academic year groups.

Table 5: Robustness - Freshman Coursemate-Companymates as Peers with Average Course Grade Control

Dependent Variable: Grade Coursemate-peer group size:	Hum/SS 2-10	Hum/SS 11-30	Hum/SS 31 or more	Math/Sci 2-10	Math/Sci 11-30	Math/Sci 31 or more
Female	0.116***	0.106***	-0.0454*	-0.0189	-0.145***	-0.101***
	(0.0245)	(0.0160)	(0.0180)	(0.0193)	(0.0240)	(0.0207)
Race/ethnicity (ref.: white and other):						
Black	-0.207***	-0.262***	-0.325***	-0.206***	-0.336***	-0.268***
	(0.0375)	(0.0264)	(0.0263)	(0.0320)	(0.0301)	(0.0330)
Asian	-0.0635	-0.0906***	-0.0998**	-0.0184	-0.136***	-0.103**
	(0.0475)	(0.0265)	(0.0304)	(0.0337)	(0.0338)	(0.0384)
Hispanic	-0.0609	-0.131***	-0.173***	-0.166***	-0.229***	-0.211***
	(0.0345)	(0.0218)	(0.0227)	(0.0293)	(0.0289)	(0.0284)
Recruited athlete	-0.0902***	-0.0689***	-0.0928***	-0.0564**	-0.153***	-0.189***
	(0.0206)	(0.0127)	(0.0140)	(0.0180)	(0.0171)	(0.0183)
Prior enlisted	0.0538	-0.0101	0.00743	0.142***	0.176***	0.117**
	(0.0379)	(0.0221)	(0.0292)	(0.0326)	(0.0280)	(0.0377)
Feeder source (ref.: Direct Entry):						
NAPS	-0.105***	-0.157***	-0.201***	0.0468	0.117***	0.271***
	(0.0304)	(0.0176)	(0.0201)	(0.0266)	(0.0232)	(0.0257)
Foundation school	-0.0766*	-0.0597**	-0.0500*	-0.115***	-0.0955**	0.0190
	(0.0383)	(0.0213)	(0.0232)	(0.0307)	(0.0301)	(0.0315)
Other	-0.0180	0.0197	-0.0208	0.229**	0.136*	0.400***
	(0.0948)	(0.0542)	(0.0552)	(0.0699)	(0.0665)	(0.0602)
Own SAT math	0.000588***	0.000384***	0.000614***	0.00338***	0.00482***	0.00613***
	(0.000177)	(0.000108)	(0.000115)	(0.000161)	(0.000169)	(0.000146)
Own SAT verbal	0.00210***	0.00281***	0.00281***	-0.000342**	0.000388**	0.000948***
	(0.000171)	(0.000106)	(0.000105)	(0.000123)	(0.000134)	(0.000126)
Peer Effects:						
Coursemate-peers' SAT math (avg.)	0.0000686	-0.00145*	-0.0000865	0.00130***	-0.00237***	-0.000681
	(0.000209)	(0.000608)	(0.000769)	(0.000192)	(0.000566)	(0.000907)
Coursemate-peers' SAT verbal (avg.)	0.00153***	-0.000798	-0.000784	0.0000993	-0.000921	-0.00173*
	(0.000183)	(0.000495)	(0.000634)	(0.000198)	(0.000597)	(0.000800)
Avg. course grade (previous year)	0.395***	0.611***	0.466***	0.522***	0.765***	Dropped
	(0.0415)	(0.0757)	(0.0800)	(0.0274)	(0.0368)	
Constant	0.0153	2.137***	1.141*	-0.595***	1.149**	-0.893
	(0.154)	(0.408)	(0.579)	(0.133)	(0.386)	(0.700)
Number of Observations	5876	20098	20148	18352	17761	13677
R^2	0.318	0.141	0.164	0.166	0.159	0.213

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. OLS estimations are carried out separately for humanities and social science course grades (columns 1-3) and math and science course grades (columns 4-6). Estimations are further stratified by the size of the coursemate-peer group associated with each grade observation (sizes 2-10, 11-30, and over 30). Coefficients for academic year dummy variables are included in the estimations but are not shown. Standard errors are clustered by company-academic year groups. Average course grade from the previous year was dropped in the far-right column due to high collinearity with academic year dummies in that subsample.